

基于独立区域划分和压缩感知的数据融合方法

张素智, 杨 芮, 陈小妮

(郑州轻工业学院 计算机与通信工程学院, 河南 郑州 450001)

摘要:数据融合是将传感器中的信息按一定准则进行综合整理,从而获得对目标的一致性描述。压缩感知(compressed sensing, CS)技术能利用更少的数据和合适的重构方法得到更精确的原始信号。针对传统数据融合方法不能有效、精确地处理海量的数据,导致融合效果不理想的问题,为提高数据融合的效率 and 融合效果,根据压缩感知理论的特点,提出了一种基于独立区域划分和压缩感知的数据融合方法。该方法运用压缩感知理论对数据进行采样以获得测量值,并通过独立数据区域划分和负载均衡方法对样本数据进行划分从而形成联合区域。计算了互信息融合权重系数,根据压缩感知系数重构方法获取融合后的数据。仿真实验结果表明,对比传统的数据融合方法,该方法具有较好的稳定性和融合效果。

关键词:数据融合;压缩感知;区域划分;负载均衡;信息熵

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2019)08-0063-04

doi:10.3969/j.issn.1673-629X.2019.08.012

Data Fusion Method Based on Independent Region Division and Compressed Sensing

ZHANG Su-zhi, YANG Rui, CHEN Xiao-ni

(School of Computer and Communication Engineering, Zhengzhou University of Light Industry,
Zhengzhou 450001, China)

Abstract: Data fusion is to synthesize the information in the sensor according to certain criteria, so as to obtain a consistent description of the target. Compressed sensing (CS) technology can use less data and appropriate reconstruction methods to get more accurate original signals. Traditional data fusion methods cannot effectively and accurately process massive amounts of data, resulting in unsatisfactory fusion effect. In order to improve the efficiency and effect of data fusion, a data fusion method based on independent data region division and compression sensing is proposed according to compression sensing theory. The method uses the CS theory to sample the data to obtain the measured values, and divides the sample data by independent data region partitioning and load balancing method to form a joint region. The mutual information fusion weight coefficient is calculated, and the fused data is obtained according to the compressed sensing coefficient reconstruction method. The simulation shows that compared with the traditional data fusion method, the proposed method has better stability and fusion effect.

Key words: data fusion; compressed sensing; region division; load balancing; information entropy

0 引言

随着信息技术的迅速发展,大数据、云计算、深度学习等新技术相继出现,人们对信息数据的需求量与日俱增,数据维度和样本规模也在逐渐增加。在进行数据处理分析时,仅面向单一的数据是远远不够的,需融合多层次多领域的数据来扩大数据处理的范围。如何对这些高维数据进行处理^[1],从而降低数据量以获取价值信息,是亟待解决的问题。由于各类数字化的

信息数据量非常大,未经过压缩和预处理的数据难以直接使用,因此,数据融合及压缩技术成为当前研究的重点。

数据融合是一种新兴信息处理技术^[2],通过基于特定规则进行优化整理、综合分析以完成所需的估计和决策。数据融合通过采集并融合以数据、信号、图像等多种形式存在于传感器中的信息,得到比各个单独的输入数据更多的信息。近年来,压缩感知理论^[3]逐

收稿日期:2018-09-02

修回日期:2019-01-03

网络出版时间:2019-03-21

基金项目:国家自然科学基金(61672470);北京市重点实验室开放课题(BKBD-2017KF08)

作者简介:张素智(1965-),男,博士,教授,CCF会员(07791M),研究方向为Web数据库、分布式计算和异构系统集成;杨 芮(1994-),女,硕士,CCF会员(75569G),研究方向为分布式并行处理及数据挖掘。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190321.0942.078.html>

渐成为数据融合中的研究热点。文献[4]对压缩感知和分布式压缩感知(DCS)的有效性进行验证,证明CS和DCS方法比无线传感器网络中的变换编码及基于模型的自适应传感提供更高的能效。文献[5]表明压缩感知方法能有效减小网络传输数据量并加快传输速度,与常规方法DANP(数据采集和无处理方法)和DATC(数据采集和变换编码方法)相比,采用压缩感知能显著地降低网络能耗并延长稀疏信号的网络寿命。文献[6]基于贝叶斯推理提出状态监测的两阶段数据融合方法,通过将来自单独状态监测系统的数据组合起来用于工业机器,可提高整个机器健康评估的可靠性。通过将CS理论应用于数据融合,可以有效地改进传统信息处理的局限性。如文献[7]基于压缩感知理论提出采样点少且结构简单的图像融合方法,可在未获取全部采样点的情况下进行融合,降低了计算复杂度,且明显使得融合的图像质量更好。文献[8]提出基于压缩感知和活动轮廓的多焦点图像融合方法,对比传统的CS图像融合方法,该方法在主观视觉和客观评价指标上有更好的融合效果。但其对于不同压缩采样矩阵的应用研究较少,且融合规则应针对测量值的特点进行有效设计。

文中提出一种基于独立区域划分和压缩感知的数据融合方法。首先通过压缩感知进行数据采样以提高采样效率,并对数据集按规则进行区域划分和负载均衡计算,对划分后的数据集采用基于信息熵的压缩感知系数融合方法。实验表明该方法降低了算法的计算复杂度,具有较好的稳定性和融合效果。

1 CS 数据表示

压缩感知理论旨在从数据本身的特性出发,找寻并去除数据中隐含的冗余度,从而达到压缩的目的。在传统信号处理中,压缩感知要求传感器以远低于Nyquist的速率^[9]进行信号采样。一般来说,如果信号是可压缩或稀疏的,即可用与变换基无关的观测矩阵将变换所得的高维信号映射到低维空间上,之后采用求解优化方法从投影中以较高的概率对原始信号进行重构。这些理论的创新,有效地推动了基于压缩感知的数据融合方法的发展。

对于长度为 N 的离散信号 $f(f \in R^N)$,其变换域为 $\Psi_i(i=1,2,\dots,N)$,对信号变换可得:

$$f = \sum_{i=1}^N x_i \Psi_i \text{ (或 } f = \Psi x \text{)} \quad (1)$$

其中, f 表示时域中的信号; Ψ 是 $N \times N$ 维稀疏矩阵; x 是表示变换系数的向量。如果 x 仅有 K 个非零值(其中 $N \gg K$)或在排列顺序后按指数级递减且逐步接近于零,则称 f 在 Ψ 上 K 稀疏。

在数据处理过程中,CS的先验要求为信号是否能稀疏表示。若信号能被压缩,并可由测量值 $y \in R^M$ 重构,则:

$$y = \Phi f = \Phi \Psi x \quad (2)$$

其中, $\Phi \{ \Phi_i, i=1,2,\dots,M \}$ 是 $M \times N$ 维($M < N$)的高斯随机观测矩阵^[10],能将 N 维信号 f 映射至 M 维空间中。通过求解优化问题^[11]可对稀疏信号 x' 重构:

$$x' = \operatorname{argmin} \|x\|_{l_1} \quad (3)$$

最后,通过 x' 求初始信号 f :

$$f = \Psi x' \quad (4)$$

2 基于独立数据区域划分的数据融合方法

压缩感知理论常用在信号的采集及重构上,然而在面对大规模数据集时,运算量较大,难以处理。因此采用分区域压缩感知技术,先对数据集进行分区划分处理,然后对每个子区域数据集分别进行观测与重构,再采用重构算法整合恢复后的各个子区域。

2.1 数据采样

无线传感网络一般由大量的传感器节点组成,这些节点具有信息采集、处理和控制的能力^[12]。假设某无线传感器网络拥有 N 个节点,每个节点在一周期内采集的数据为 $t_i, i=1,2,\dots,n$ 。整个无线传感网络数据可构成一个矢量,即

$$T = [t_1, t_2, \dots, t_n]^T \quad (5)$$

如果需要得到完整的信息,应采集信号 T 的 N 个采样的样本,并通过压缩感知方法获取的信号变换系数 $\alpha(\|\alpha\| \ll N)$ 来恢复完整的无线传感信号。

由于 T 通常较大,可能包含数千个无线传感的节点数据,采用压缩感知能有效减少信息采集量。对一个信号 T ,若存在变换域 Ψ ,使得 T 能在 Ψ 上 q 稀疏表示,则

$$\Psi = [\psi_1, \psi_2, \dots, \psi_q]^T \quad (6)$$

无线传感网络的数据采样表示为:

$$T = \Psi P = \sum_{i=1}^N \psi_i P_i \quad (7)$$

其中, P 为 T 的稀疏表示。

压缩感知利用高斯随机采样矩阵 Φ ,对传感节点上的数据进行采样,综合上述,经压缩采样取得的测量值为:

$$y_i = \phi T_i, i=1,2,\dots \quad (8)$$

其中, $\phi = \{ \phi_{j,i} \}$ 为采样矩阵,矩阵中的元素满足独立同分布的特性,其方差为 $1/M$ 。

2.2 非规则区域数据划分

2.2.1 独立区域划分

按照划分准则进行区域划分,将采样数据动态分组多个独立的相关自包含数据区域,以消除数据耦

合性,使得同一区域中的数据具有较高的相似度,不同区域中的数据间不相关。根据近邻数据思想,按规则聚类,形成近邻动态区域。

(1)对于待聚类的点迹集合 $U = \{U_1, U_2, \dots, U_n\}$, 元素两两之间距离为:

$$d_{ij} = d(u_i, u_j) = \|u_i - u_j\|, i, j = 1, 2, \dots, n \quad (9)$$

且 $i \neq j$

(2)在计算过程中,从 U 中找出与 u_i 距离小于阈值 T 的各元素形成初始簇 C_i :

$$C_i(u_i) = \{u_m \mid d_{im} < T, d_{im} \in D_i\}, u_m \in U \quad (10)$$

其中, T 为提前设定的距离阈值,文中设为 20。则聚类划分为 $C(u) = \{C_i(u_i) \mid i = 1, 2, \dots, n\}$, 其划分数目为 n 个。

2.2.2 负载均衡

根据独立区域划分得到的分区大小通常不相等。因此,为实现负载均衡^[13],对聚类结果 $C(u)$, 将其大致分为 d 份,分别记为 $P_c, P = \{P_c \mid c = 1, 2, \dots, d\}$, 并行节点数 k , 采用基于区域计算负荷的均衡划分方法。

(1)各节点的计算量为:

$$\text{avg}L = \frac{1}{k} \sum_{i=1}^d |P_i| \quad (11)$$

(2)按照分区大小进行排序,得到有序集合 $P' = \{P'_i \mid |p'_1| \geq |p'_2| \geq \dots \geq |p'_d|\}$ 。

(3)依次从有序集合 P' 的两端取元素,组成 k 个子集 $M = \{M_i \mid i = 1, 2, \dots, k\}$, 满足:

$$|M_1| \approx |M_2| \approx \dots \approx |M_k|, P' = \bigcup_{i=1}^k M_i, M_i \cap M_j = \emptyset, \forall i, j \quad (12)$$

文中将样本数据进行近邻数据区域划分,并对划分后的各区域子集设置标号矩阵为 $\{A_1, A_2, \dots, A_k\}$ 。假设两个区域集的标号矩阵分别为 A_1 和 A_2 , 则联合矩阵为:

$$A = NA_1 + A_2 \quad (13)$$

其中, $N = N_1 N_2$, N_1 和 N_2 分别表示 2 个样本数据集的划分数量, N 表示 2 个样本数据集划分的所有可能,因此 N 为一个足够大的整数。联合区域矩阵可以保证不同的数据集具有相同的划分标准。

2.3 数据融合

通过划分,得到的划分集为 $M_i (i = 1, 2, \dots, k)$, 为了更加精确地得到数据轨迹,需要对数据进行融合。采用基于信息熵^[14]的 CS 系数融合方法,计算两个类的信息熵:

$$E(M) = - \sum_k p_k \log_2 p_k \quad (14)$$

其中, p_k 为随机变量 M 在信号中出现的概率。

联合信息熵为:

$$E(M_1, M_2) = - \sum_{M_1} \sum_{M_2} P(M_1, M_2) \log_2 P(M_1, M_2) \quad (15)$$

互信息(MI)为:

$$I(M_1, M_2) = \sum_{M_1} \sum_{M_2} P(M_1, M_2) \log_2 \frac{P(M_1, M_2)}{P(M_1)P(M_2)} \quad (16)$$

对两个类 M_1, M_2 , 其各自的权重计算系数为:

$$\omega_1 = \frac{E(M_1)}{E(M_1, M_2)} - \frac{I(M_1, M_2)}{2E(M_1)} \times \frac{I(M_1, M_2)}{E(M_1, M_2)} \quad (17)$$

$$\omega_2 = \frac{E(M_2)}{E(M_1, M_2)} - \frac{I(M_1, M_2)}{2E(M_2)} \times \frac{I(M_1, M_2)}{E(M_1, M_2)} \quad (18)$$

根据独立区域划分,两个区域集的联合矩阵 A 的融合权重为:

$$W'_k(i, j) = \frac{\sum_{(m, n)} \delta(A(m, n) - A(i, j)) W_k(m, n)}{\sum_{(m, n)} \delta(A(m, n) - A(i, j))}, \quad k \in \{1, 2\} \quad (19)$$

其中, $\delta(x)$ 表示单位冲激函数,定义为:

$$\delta(t) = \begin{cases} 0 & t \neq 0 \\ 1 & t = 0 \end{cases} \quad (20)$$

假设 2 个样本数据集相应的区域集测量值分别为 y_1 和 y_2 , 则每个区域集融合后的测量值为:

$$y' = \omega_1 y_1 + \omega_2 y_2 \quad (21)$$

基于压缩感知的数据融合算法具体步骤如下:

(1)使用压缩感知方法进行数据采样。通过高斯随机观测矩阵对数据的稀疏表示形态进行测量,得到测量值 $\{y_1, y_2, \dots, y_n\}$;

(2)对采样数据集按规则进行独立区域划分,计算负载均衡,并对区域集设置联合矩阵。

对待融合的划分后数据集 $\{M_1, M_2, \dots, M_k\}$, 计算两个样本类的信息熵、联合信息熵和互信息,获取权重融合后的测量。

3 实验与分析

3.1 实验数据

为了衡量算法的优劣,采用文献[15]中的实验数据,使用三个热电偶对恒温箱进行温度检测,六次的观测值如表 1 所示。

表 1 温度传感器的观测值(温度设定 900℃)

传感器	观测次数					
	1	2	3	4	5	6
1	899.5	905.3	901.9	900.6	889.9	899.4
2	898.3	875.9	888.1	886.2	907.5	904.4
3	896.7	906.8	898.2	904.0	896.4	891.6

3.2 融合结果与分析

为了验证算法的精度,分别采用平均值方法、文献[15]中的可靠性算法、文中方法对表 1 中的观测数据

进行融合,得到的融合结果如表 2 所示。

由表 2 的融合结果可以看出,文中方法的多次观测值都较接近于实际值,且对比三种方法的绝对误差和,平均值方法为 16.433 3,可靠性算法为 16.308 9,而文中方法的误差和远低于前两种方法,为 12.222 0。因此,文中方法具有一定的优势。

表 2 三种融合方法的融合结果对比

观测次数	平均值方法		可靠性算法		文中方法	
	融合值	绝对误差	融合值	绝对误差	融合值	绝对误差
1	898.166 7	1.833 3	898.241 4	1.758 6	902.400 0	2.400 0
2	896.000 0	4.000 0	896.127 0	3.873 0	900.890 0	0.892 0
3	896.066 7	3.933 3	896.057 1	3.942 9	898.200 0	1.800 0
4	896.933 3	3.066 7	896.864 5	3.135 5	901.330 0	1.330 0
5	897.933 3	2.066 7	897.919 7	2.080 3	903.950 0	3.950 0
6	898.466 7	1.533 3	898.481 4	1.518 6	898.150 0	1.850 0
误差和	16.433 3		16.308 9		12.222 0	

为了进一步验证算法的稳定性,使用 Matlab 仿真软件模拟 10 个传感器对 900 ℃ 恒温箱随机观测 100 次的实验,并对采样获取的数据采用三种方法进行融合实验,其绝对误差对比如图 1 所示。

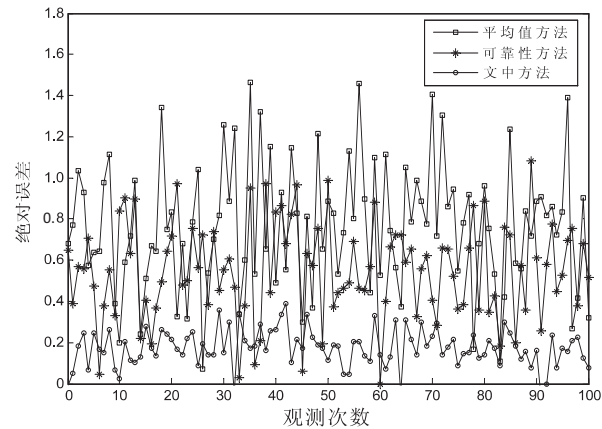


图 1 三种融合方法的绝对误差对比

由图 1 中三种融合方法的绝对误差曲线变化可知,文中方法的绝对误差曲线整体上位于平均值方法和可靠性方法曲线的下方,且误差曲线变化较为平缓,而平均值方法和可靠性方法的误差曲线波动较大,且误差值相对较高。因此,文中方法具有一定的稳定性,融合效果较好。

4 结束语

在数据融合中,压缩感知理论的出现推动了传感器中数据信号处理的发展。文中运用压缩感知理论的思想,提出基于独立区域划分和压缩感知的数据融合算法,对数据采样获取测量值后,通过独立区域划分样本集,并计算互信息融合权重,最后采用压缩感知系数重构方法获取融合后数据。

实验结果表明,文中算法具有较好的稳定性和融合效果。

参考文献:

[1] 何兴高,李蝉娟,王瑞锦,等. 基于信息熵的高维稀疏大数据降维算法研究[J]. 电子科技大学学报,2018,47(2):235-241.

[2] TAN Rui,XING Guoliang,YUAN Zhaohui,et al. System-level calibration for data fusion in wireless sensor networks[J]. ACM Transactions on Sensor Networks,2013,9(3):1-27.

[3] DONOHO D L. Compressed sensing[J]. IEEE Transactions on Information Theory,2006,52(4):1289-1306.

[4] RAZZAQUE M A,DOBSON S. Energy-efficient sensing in wireless sensor networks using compressed sensing[J]. Sensors,2014,14(2):2822-2859.

[5] KARAKUS C,GURBUZ A C,TAVLI B. Analysis of energy efficiency of compressive sensing in wireless sensor networks [J]. IEEE Sensors Journal,2013,13(5):1999-2008.

[6] JARAMILLO V H,OTTEWILL J R,DUDEK R,et al. Condition monitoring of distributed systems using two-stage Bayesian inference data fusion[J]. Mechanical Systems & Signal Processing,2017,87:91-110.

[7] 张 伟,曾凡仔,曾庆光. 基于压缩感知理论的图像融合方法[J]. 计算机工程与应用,2012,48(12):194-197.

[8] 陈 畅,覃 征. 基于压缩感知和活动轮廓的多焦点图像融合[J]. 清华大学学报:自然科学版,2014,54(12):1539-1543.

[9] MASSA A,ROCCA P,OLIVERI G. Compressive sensing in electromagnetics-a review[J]. IEEE Antennas & Propagation Magazine,2015,57(1):224-238.

[10] 王 侠,王 开,王青云,等. 压缩感知中的确定性随机观测矩阵构造[J]. 信号处理,2014,30(4):436-442.

[11] 任越美,张艳宁,李 映. 压缩感知及其图像处理应用研究进展与展望[J]. 自动化学报,2014,40(8):1563-1575.

[12] 陈善雄,何中市,熊海灵,等. 一种基于压缩感知的无线传感信号重构算法[J]. 计算机学报,2015,38(3):614-624.

[13] 苏金树,郭文忠,余朝龙,等. 负载均衡感知的无线传感器网络容错分簇算法[J]. 计算机学报,2014,37(2):445-456.

[14] 赵兴旺,梁吉业. 一种基于信息熵的混合数据属性加权聚类算法[J]. 计算机研究与发展,2016,53(5):1018-1028.

[15] 孙 勇,景 博. 基于支持度的多传感器一致可靠性融合[J]. 传感技术学报,2005,18(3):537-539.